

Ajuda:Expressões regulares

Origem: Wikipédia, a enciclopédia livre.



A informação presente nesta página contém provavelmente erros que devem ser corrigidos.

Isso deve-se às diferentes interpretações das expressões regulares conforme a extensão do MediaWiki ou programa usado.

Esta página descreve caracteres especiais, também chamados metacaracteres (assim como sequências destes) usados em expressões regulares que podem ou não ser utilizados nas seguintes páginas:

- MediaWiki:Spam-blacklist (mw:Extension:SpamBlacklist)
- MediaWiki:Spam-whitelist (idem)
- MediaWiki:Titleblacklist (mw:Extension:Title Blacklist)
- MediaWiki:Titlewhitelist (idem)
- MediaWiki:Bad image list (mw:Extension:Bad Image List)
- MediaWiki:Captcha-addurl-whitelist (mw:Extension:ConfirmEdit)
- Especial:AbuseFilter (Wikipedia:Filtro de edições)
- Usuário:Salebot/Config (robô anti-vandalismo)

Outros programas/recursos não abordados especificamente nesta página (mas a mesma pode servir de ajuda):

- AutoWikiBrowser - Recomenda-se a consulta de en:Wikipedia:AutoWikiBrowser/Regular expression (em inglês).
- WikiEd usa a sintaxe de expressões regulares JavaScript disponíveis em

Guia de edição

Criar uma conta
 Menus e ferramentas
 Como começar uma página
 Formatar artigos
 Como usar imagens, áudios e vídeos
 Ligar artigos entre si
 Fazer notas de rodapé
 Ligar artigos a projetos relacionados
 Como adicionar ligações externas
 Categorizar
 Ligar artigos a outras Wikipédias
 Usar predefinições
 Criar predefinições
 Usar o acordo ortográfico
 Usar tabelas
 Usar fórmulas matemáticas
 Renomear uma página
 Reverter edições
 Fundir artigos
 Tamanho de páginas
 Dividir artigos
 Redirecionar páginas
 Desambiguar termos
 Traduzir artigos
 Assinar mensagens
 Arquivar discussões
 Dicas e truques

Guia de edição avançado

HTML
CSS
Funções sintáticas
Palavras mágicas
JavaScript
MediaWiki
Expressões regulares
Extensões
Desenvolvedores
Outros guias
Portal de ajuda
Primeiros passos
FAQ de edição
Livro de estilo
Guia de consulta e reprodução
Guia de eliminação
editar (http://pt.wikipedia.org/w/index.php?title=Predefinição:Guia_de_edição&action=edit)

http://developer.mozilla.org/en/Core_JavaScript_1.5_Guide/Regular_Expressions (em inglês)

- Wikipédia:Software/Anti-vandal tool/badwords
- Outras páginas do MediaWiki como Especial:Pesquisar_links e alguns Gadgets normalmente usam muito poucos caracteres especiais (como, por exemplo, apenas *) e as informações sobre as mesmas estão normalmente nas respetivas páginas.

Nota: nem todos os caracteres especiais se podem usar em qualquer página (ou programa) descrita anteriormente, por isso esta página contém secções de exemplos conforme a página (ver #Exemplos práticos) e a indicação em que páginas funciona na tabela que se segue nas colunas "Spam", "Title", "Salebot" e "Filtro". A sintaxe utilizada pelas extensões do MediaWiki é a PCRE.

Índice

- 1 Descrição dos caracteres especiais
- 2 Caractere de escape
- 3 Exemplos práticos
 - 3.1 MediaWiki:Titleblacklist
 - 3.2 MediaWiki:Spam-blacklist

- 3.3 MediaWiki:Spam-whitelist
- 4 Lista de nomes de classes Unicode
 - 4.1 Propriedades
 - 4.2 Sistema de escrita
 - 4.3 Blocos
- 5 Testar expressões regulares
 - 5.1 Sítios (on-line)
 - 5.2 Programas
- 6 Notas
- 7 Ver também
- 8 Ligações externas

Descrição dos caracteres especiais

Caractere especial ou sequência de caracteres especiais	Descrição	Spam	Title	Salebot	Filtro
	A barra vertical significa "ou"; detecta o termo anterior ou posterior. Pode ser combinado com várias barras verticais. A utilização sem ser dentro de parêntesis pode depender do programa/extensão em que é utilizado, ver exemplos seguintes para saber se se pode utilizar.		✓	✓	✓
<i>sim não talvez</i>	Neste exemplo deteta "sim", "não" e "talvez"				✓
<i>muit(o as)</i>	Neste caso detecta "muito" e "muitas"			✓	✓
.	qualquer caractere (inclui espaços)	✓	✓	✓	
*	(0, 1 ou mais) nenhuma, uma ou mais repetições do termo anterior	✓	✓	✓	
.*	Repetição (ou não) de qualquer caractere (inclui espaços) uma ou mais vezes.		✓	✓	
(ha)*	Repetição (ou não) de "ha" uma ou mais vezes. Deteta "ha", "haha", "hahaha", etc.		✓	✓	
\d*	Repetição (ou não) de um dígito uma ou mais vezes. Deteta "3", "823", "5961837254", etc.		✓	✓	

alt\w*	Repetição (ou não) de um caractere alfanumérico uma ou mais vezes. Deteta "alto", "alternativa", "super!altamente", etc. mas não quando "alt" é seguido de um dígito como "super!alt1".		✓	✓	
+	(1 ou mais) uma ou mais repetições do termo anterior. Equivalente a {1, }		✓	✓	
?	(0 ou 1) Nenhuma ocorrência ou uma do termo anterior, ou seja, o termo anterior é opcional (mas tenta encontrar primeiro uma ocorrência e só depois a ausência). Caso o ? seja colocado após um dos símbolos ? * + ou da expressão { <i>mínimo</i> , <i>máximo</i> } assume primeiro a ausência e só depois a ocorrência.		✓	✓	
\	Este caractere especial tem duas funções conforme o que lhe segue: <ul style="list-style-type: none"> é utilizado como caractere de escape, para que se interprete um caractere especial como caractere normal. Por exemplo, a expressão quem\? deteta "quem?" e não "que" e "quem" caso se usasse quem?. é utilizado para uma sequência especial como por exemplo número \d que deteta "número 8", "número 17456", etc. e não "número d" pois o "d" não é um caractere especial para que o "\" seja usado para caractere de escape. 	✓	✓	✓	✓
\A					
\b	Restringe ao início, o meio ou o fim da expressão que antecede ou precede. Por exemplo \bsim deteta "é simplesmente" mas não deteta "assim".	✓	✓	✓	
\B	Qualquer posição do texto desde que não seja no início ou no fim de uma palavra. Por exemplo \Btur\B deteta "altura" mas não deteta "altur". Já \btur\B deteta "turismo" mas não "altura"				
\d	qualquer dígito (0 a 9)		✓	✓	
\D	qualquer caractere que não seja um dígito				
\G					
\s	Espaço em branco		✓	✓	✓
\S	Tudo excepto o espaço			✓	
\w	qualquer caractere alfanumérico incluindo traço inferior; é o equivalente a [a-z0-9_]		✓	✓	
\W	qualquer caractere que não seja alfanumérico, incluindo traço inferior; é o equivalente a [^a-z0-9_]		✓	✓	
\Z					

<code>\z</code>				
<code>\0</code>		✓		
<code>\1</code>	Pode-se usar <code>\1 \2 \3 \4 \5 \6 \7 \8 \9</code> para capturar um grupo anterior (dentro de parêntesis). Exemplo <code>(ha)\1{2,}</code> permite detectar "hahaha". Exemplo <code>.*(\.)\1{10}.*</code> permite detectar 10 caracteres iguais repetidos. Outro exemplo <code>dois (tigres esquilos) três \1</code> deteta "dois tigres três tigres" e "dois esquilos três esquilos", mas não "dois tigres três esquilos".	✓	✓	
<code>()</code>	Permite definir um grupo de expressões ou caracteres. Captura também a expressão para referência posterior (ver expressão anterior <code>\1</code>), a não ser que se use como <code>(?:)</code> que não captura. Os parêntesis também permitem aplicar expressões regulares posteriormente a esse grupo (e não apenas a um caractere) como por exemplo <code>(ha)+</code>	✓	✓	✓
<code>(forum site) novo</code>	Neste caso usa um elemento comum "novo" para detetar várias expressões como "fórum novo" e "site novo"		✓	✓
<code>(bom){2}</code>	Neste caso agrupa "bom" como um só elemento para considerar a repetição que detecta "bombom"		✓	
<code>(muito)? bom</code>	Neste caso agrupa "bom" como um só elemento para considerar a ocorrência ou não desse elemento, detentando "muito bom" e "bom"		✓	
<code>[]</code>	Permite especificar os caracteres a detetar	✓	✓	✓
<code>[abc]</code>	Deteta "a", "b" ou "c"	✓	✓	✓
<code>[0-4]</code>	Exemplo para detetar um intervalo de caracteres. Neste caso deteta "1", "2", "3" e "4"	✓	✓	✓
<code>[a-d]</code>	Deteta "a", "b", "c" e "d"	✓	✓	✓
<code>[0-4a-d]</code>	Deteta "1", "2", "3", "4", "a", "b", "c" e "d"	✓	✓	
<code>[a-g-[b-e]]</code>	Exemplo para subtração de um conjunto. Neste exemplo significa detetar o intervalo de letras de "a" a "g" excepto o intervalo de letras de "b" a "e". Deteta então "a", "f" e "g". Em princípio esta expressão não funciona.			
<code>[a-z-[aeiou]]</code>	Neste caso deteta todas as consoantes (pode ser interpretar a expressão como <i>alfabeto exceto vogais</i>). Em princípio esta expressão não funciona.			
<code>[a-z&&[bdg]]</code>	Neste caso deteta todas as letras do alfabeto exceto as letras "b", "d" e "g". Funciona como interseção.			
	Caso o hífen esteja logo no início é interpretado como tal. No exemplo corresponde ao "hífen", letra "a",			

<code>[-abc]</code>	"b" e "c"		✓		
<code>[^]</code>	Permite especificar os caracteres a não detetar (negação). Ver também outro uso do caractere ^		✓	✓	
<code>[^abc]</code>	qualquer caractere que não seja "a", "b" ou "c"		✓	✓	
<code>[^a-f]</code>	qualquer caractere que não seja uma letra do alfabeto entre "a" e "f"; o mesmo que <code>[^abcdef]</code>		✓	✓	
<code>[^0-4]</code>	qualquer caractere que não seja um número entre "0" e "4"; o mesmo que <code>[^01234]</code>		✓	✓	
<code>{ }</code>	Quantificador. Permite especificar o número de vezes que a expressão anterior ocorre, quer o número exacto, quer o mínimo, quer o mínimo e máximo.		✓	✓	
<code>{ número }</code>	Número exacto de vezes que o elemento anterior se repete		✓	✓	
<code>{ mínimo , }</code>	Número mínimo de vezes que o elemento anterior se repete		✓	✓	
<code>{ mínimo , máximo }</code>	Número mínimo e máximo de vezes que o elemento anterior se repete		✓	✓	
<code>(? : expressão)</code>	Permite definir um grupo de expressões ou caracteres. Não captura o grupo para referência posterior ao contrário de <code>()</code>	✓	✓		
<code>(? ! expressão)</code>	Grupo em negação não capturado. Por exemplo <code>muito (? ! mais menos)</code> deteta tudo o que tiver "muito" e não seja seguido pelas palavras "mais" ou "menos" como "muito bom", "muito qualquer", etc.				
<code>(? = expressão)</code>					
<code>(? < ! expressão)</code>			✓		
<code>(? < = expressão)</code>					
<code>(? > expressão)</code>					
<code>\ 045</code>	Código ASCII no sistema octal. Neste exemplo "045" corresponde ao símbolo "%".				
<code>\ x25</code>	Código hexadecimal. Neste exemplo "25" corresponde ao símbolo "%".				
<code>\ x</code>					
<code>\ x { 25 }</code>	Semelhante ao anterior mas na sintaxe PCRE para código hexadecimal. Neste exemplo "25" corresponde também ao símbolo "%".		✓		
<code>\ u0025</code>	Código Unicode. Neste exemplo "0025" corresponde ao símbolo "%". Em princípio não é possível usar esta expressão nas páginas do MediaWiki. Para isso use o <code>\ x { 25 }</code>		✗?		

<code>\p{nome da classe}</code>	Caracteres de um determinada classe. Por exemplo, <code>\p{IsHebrew}</code> em que "IsHebrew" é o nome da classe Unicode para as letras da língua hebraica. Para uma lista completa ver secção desta página #Lista de nomes de classes Unicode		✓		
<code>\P{nome da classe}</code>	Caracteres que não sejam da classe definida. É semelhante ao anterior mas em negação.		✓		
<code>\P{^nome da classe}</code>	Caracteres que não sejam da classe definida. É semelhante a <code>\P{nome da classe}</code> (notar que o "P" está em maiúscula")		✓		
<code># comentárioXPTO</code>	Utilizado para comentários. Todo o texto que se encontre após # é interpretado como comentário	✓	✓	✓	
<code>(?#comentárioXPTO)</code>	Utilizado para comentários, normalmente dentro das próprias expressões regulares. Todo o texto que se encontre entre os parênteses não é considerado uma expressão regular. Por exemplo <code>muitos?(?#comentárioXPTO)tigre</code> deteta "muitostigre" e "muitotigre"				
<code>^</code>	Início do texto. Por exemplo <code>^34</code> deteta "34" num texto que comece com "34 hs gj ada hs". Não confundir com <code>\b</code> que tem uma funcionalidade diferente.		✗ ¹	✓	
<code>\$</code>	Fim do texto. Por exemplo <code>34\$</code> deteta "34" num texto que termine com "hs gj ada hs34". Não confundir com <code>\b</code> que tem uma funcionalidade diferente.		✗ ¹		
<code>(?i:expressão)</code>	Utilizado para considerar maiúsculas e minúsculas. Por exemplo, <code>(?i:Meta)</code> corresponde a "Meta", "META", "meta", "MeTa", etc.		✗ ²		
<code>(?-i:expressão)</code>	Inverso do anterior. Por exemplo <code>(?i:sím) (?-i:bolo)</code> deteta "SIMbolo" e "símbolo" mas não "símBOLO"		✗		
<code>\Qexpressão\E</code>	Interpreta literalmente a expressão como caracteres.				
<code>\Q+({\E</code>	Interpreta literalmente a expressão como caracteres. Deteta "+{"				
<code>(expressão) (? (grupo) sim não)</code>					
<code>(expressão) (? (grupo) sim)</code>					

Caractere de escape

Quando se quer que o programa que usa expressões regulares interprete um caractere especial como um caractere normal (ou literal) é necessário usar um caractere de escape `\` antes de cada caractere especial. Por exemplo, para se detetar o texto **muito + que** é necessário usar a expressão `muito \+ que\?` para que se interprete literalmente como texto e não expressão regular. Notar que para cada caractere especial é necessário usar o caractere de escape, mesmo que estejam seguidos. Por exemplo, em `+?*` deve-se usar `\+?*` para interpretar como caracteres normais.

- nas páginas MediaWiki:Titleblacklist e MediaWiki:Spam-blacklist os símbolos `^ \ * . () { } [] $ + - | ? < >` para serem interpretados literalmente têm de ser colocados após `\` Por exemplo, para bloquear o termo exato **muito+mais** deve-se usar a expressão `muito\+mais\\` de forma a que o símbolo `+` e `\` não sejam interpretados como uma expressão regular mas sim como caracteres. Isto não se aplica quando os símbolos `\ * . + ?` estão imediatamente a seguir a `[` ou imediatamente antes de `]` como por exemplo `[-!€]`
- na página MediaWiki:Titleblacklist o símbolo `/` tem de ser colocado da seguinte forma `\/`

Exemplos práticos

Notas:

- Nos exemplos mostrados não se considera a diferença entre maiúsculas e minúsculas, pois o MediaWiki usa por defeito essas variações. Ou seja, a expressão **raios** bloqueia "Raios", "raios", "RAIOS", "RaIoS", etc.
- Nos exemplos utiliza-se muito as expressões entre `\b` para mostrar apenas alguns exemplos na coluna "Palavras detectadas"

MediaWiki:Titleblacklist

Exemplo	deteta	não deteta	Explicação
viado	"viado" "viados" "aviado" "aviador" "um viado de", etc.	"vlado", etc	bloqueia todas as palavras que começam, terminem ou tenham no meio a expressão "viado". Notar que neste caso bloqueia também expressões válidas como "aviador". Deve-se ter algum cuidado com este caso pois pode bloquear muitas palavras válidas, principalmente quando a palavra/expressão a bloquear for muito curta. A título de exemplo vi iria bloquear ainda mais páginas ao contrário de montes de viados que não originaria problemas pois provavelmente não deve haver um título de uma página ou conta de usuário válida com essas palavras. Teoricamente (a confirmar) obtêm-se os mesmos resultados com <code>.*viado.*</code>
	"viado" "viados"	"aviado"	Exemplo semelhante ao anterior, mas que bloqueia apenas as palavras que começam com

<code>\bviado</code>	"viadores" "viadonte", etc.	"mesmoviado", etc	viado
<code>viado\b</code>	"viado" "aviado" "wq12wprviado", etc.	"viados", etc	Bloqueia apenas as palavras que terminam com viado
<code>\bviado\b</code>	"viado"	"umviado" "viadão"	Bloqueia apenas a palavra viado
<code>p.ta</code>	"pata" "pota" "aap1taa" "computador", etc.	"pta", etc	Semelhante ao anterior, aceita qualquer caractere (apenas 1) entre p e ta . Este exemplo nunca deve ser utilizado desta forma pois deteta demasiadas palavras válidas.
<code>puta.*pariu</code>	"puta que pariu" "puta pariu" "aputa de 1parius sem", "computador que comparium" etc.	"pariu puta" "putapariu", etc	Detecta a palavra "puta" e "pariu" quando estas estão seguidas. Tel algum cuidado com este exemplo, pode eventualmente detectar palavras corretas.
<code>\bputa\b.*\bpariu\b</code>	"puta que pariu" "puta pariu", etc	"putapariu" "umputa pariu", etc	Semelhante ao anterior mas só detecta as palavras "puta" e "pariu" quando estão isoladas
<code>\bmi?erda\b</code>	"mierda" "merda"	"m3rda" "merdices" "umamerda", etc	O ? neste caso indica que a letra anterior i é opcional.
<code>\bfix(e ola ix)\b</code>	"fixe" "fixola" "fixix"	"fixone" "fix3" "fix", etc	Utilizado para palavras com variantes mas com uma parte em comum.
<code>\bbarda[]?merda\b</code>	"barda merda" "bardamerda"	"barda-merda" "barda merda", etc	Para tornar os espaços opcionais, estes têm de ser colocados entre colchetes.

<code>[0-9]º ano</code>	"meu 1º ano de" "teu 10º anode", etc.	"9ºano" "9o ano", etc	Caso especial que considera os algarismos entre 0 e 9. É uma forma simplificada de obter o mesmo resultado com <code>[0123456789]º ano</code> ou <code>(0 1 2 3 4 5 6 7 8 9)º ano</code>
<code>\b1?[0-9]º ano\b</code>	"meu 1º ano de" "teu 10º ano de", etc.	"meu 1º ano de" "10º anode", etc.	

MediaWiki:Spam-blacklist

Exemplo	Endereços detetados	Explicação
<code>\bexemplo\.com</code>	"www.exemplo.com" "sv1.exemplo.com/34/pag.html", etc.	Exemplo mais comum que bloqueia todas as páginas de um domínio.
<code>\bexemplo\.com/game/</code>	"www.exemplo.com/game/" "sv1.exemplo.com/game/pag.html", etc.	Exemplo para bloquear apenas determinados directórios de um domínio e não todas as páginas desse domínio.
<code>\bexemplo\.(:com net org pt com\.br)\b</code>	"www.exemplo.com" "www.exemplo.pt" "www.exemplo.com.br" "sv1.exemplo.pt/pag.html", etc.	Para bloquear vários domínios com o mesmo nome.
<code>\bserver\d\.exemplo\.com\b</code>	"server1.exemplo.com" "server4.exemplo.com" "server9.exemplo.com/pag.htm", etc.	Neste caso o <code>\d</code> corresponde a qualquer dígito (0 a 9).
<code>casino-online</code>	"casino-online.com" "newcasino-onlines.pt" "www.w3.com/casino-online1.htm", etc.	Bloqueia todos os URL que contenham "casino-online". Ao usar este método deve-se ter em conta que bloqueia também directórios de sites.
<code>@exemplo.com</code>	"info@exemplo.com" "qual@exemplo.com", etc.	
<code>72\.232\.68\.234</code>		
<code>\bgoogle\...\{1,5}/cse\b</code>		
<code>exemplo[0-9a-z-]*\.com</code>		
<code>www\.google\.com/search\? source=ig&hl=en&q=music\+education\+by\+color</code>		
<code>upload\.wikimedia\.org/. *XRumer.screenshot\.gif</code>		

MediaWiki:Spam-whitelist

Exemplo	Explicação
<code>\bwww\.exemplo\.org\b</code>	Desbloqueia apenas a página principal do site.
<code>\bwww\.exemplo\.org/results.html\b</code>	Desbloqueia apenas a página "www.exemplo.org/results.html" desse site.
<code>\bwww\.exemplo\.org/results/\b</code>	Desbloqueia apenas o diretório "www.exemplo.org/results/" e todas as subpáginas que estejam nesse diretório.

Lista de nomes de classes Unicode

As seguintes listas podem ser utilizadas de várias formas:

- `\p{nome da classe}`
- `\P{nome da classe}`
- `\p{^nome da classe}`
- `\P{^nome da classe}`

Também podem ser combinadas:

- `\p{L|P}` (a confirmar)
- `[\p{L} \p{P}]`

A lista seguinte "Propriedades" pode ser utilizada de duas formas: `\p{Lu}` ou `\pLu` ' ' ' com as quais se obtem o mesmo resultado.

Propriedades

Nome	Nome alternativo	Descrição
L	Letter	Todas as letras
Lu	Uppercase_Letter	Todas as letras maiúsculas

Ll	Lowercase_Letter	Todas as letras minúsculas
Lt	Titlecase_Letter	Quando a primeira letra de uma palavra está em maiúscula.
Lm	Modifier_Letter	
Lo	Other_Letter	
L&	Letter&	O mesmo que Ll, Lu, ou Lt; ou seja "Todas as letras maiúsculas", "todas as letras minúsculas" e "Quando a primeira letra de uma palavra está em maiúscula"
N	Number	Todos os números
Nd	Decimal_Digit_Number	Dígito decimal. Para <code>\p{Nd}</code> equivale a usar <code>\d</code> Para <code>\P{Nd}</code> equivale a usar <code>\D</code>
Nl	Letter_Number	
No	Other_Number	
P	Punctuation	Todos os sinais de pontuação (ver lista de alguns em Predefinição:Pontuação)
Pc	Connector_Punctuation	
Pd	Dash_Punctuation	traços (incluindo hífen)
Pi	Initial_Punctuation	
Pf	Final_Punctuation	
Ps	Open_Punctuation	Toda a pontuação de abertura como ([{
Pe	Close_Punctuation	Toda a pontuação de fecho como)] }
Po	Other_Punctuation	
S	Symbol	Símbolos como os monetários, matemáticos, etc.
Sc	Currency_Symbol	Todos os símbolos monetários como \$ € £ ¢ ¥ etc. (ver en:Currency sign)
Sk	Modifier_Symbol	
Sm	Math_Symbol	Todos os símbolos matemáticos
So	Other_Symbol	
M	Mark	Todos os diacríticos
Mn	Non_Spacing_Mark	

Mc	Spacing_Combining_Mark	
Me	Enclosing_Mark	
Z	Separator	Espaços em branco
Zs	Space_Separator	Espaço
Zl	Line_Separator	
Zp	Paragraph_Separator	
C	Other	
Cc	Control	
Cf	Format	
Cn	Unassigned	
Co	Private_Use	
Cs	Surrogate	

Sistema de escrita

Por sistema de escrita entenda-se os caracteres utilizados numa determinada língua. É diferente dos blocos descritos na tabela seguinte na medida em que abarca símbolos de vários blocos Unicode. Por exemplo **Latin** inclui os blocos Unicode:

- BasicLatin
- Latin-1 Supplement
- LatinExtendedA
- LatinExtendedB
- LatinExtendedAdditional
- Diacritics

Nome	Descrição
Arabic	
Armenian	en:List of Unicode characters#Armenian
Balinese	
Bengali	
Bopomofo	
Braille	
Buginese	
Buhid	
CanadianAboriginal	
Cham	
Cherokee	
Common	
Coptic	
Cuneiform	
Cypriot	
Cyrillic	Alfabeto cirílico
Deseret	
Devanagari	
Ethiopic	
Georgian	
Glagolitic	
Gothic	
Greek	
Gujarati	

Gurmukhi	
Han	
Hangul	
Hanunoo	
Hebrew	en:List of Unicode characters#Hebrew
Hiragana	
Inherited	
Kannada	
Katakana	
Kharoshthi	
Khmer	
Lao	
Latin	
Limbu	
LinearB	
Malayalam	
Mongolian	
Myanmar	
NewTaiLue	
Nko	
Ogham	
OldItalic	
OldPersian	
Oriya	

Osmanya	
PhagsPa	
Phoenician	
Runic	
Shavian	
Sinhala	
Sundanese	
SylotiNagri	
Syriac	
Tagalog	
Tagbanwa	
TaiLe	
Tamil	
Telugu	
Thaana	
Thai	
Tibetan	
Tifinagh	
Ugaritic	
Yi	

Blocos

Nome	Descrição
------	-----------

InAlphabeticPresentationForms	
InArabic	
InArabicPresentationFormsA	
InArabicPresentationFormsB	
InArabicSupplement	
InArmenian	en:List of Unicode characters#Armenian
InArrows	
InBalinese	
InBasicLatin	
InBengali	
InBlockElements	
InBopomofo	
InBopomofoExtended	
InBoxDrawing	
InBraillePatterns	
InBuginese	
InBuhid	
InCham	
InCherokee	
InCJKCompatibility	
InCJKCompatibilityForms	
InCJKCompatibilityIdeographs	
InCJKRadicalsSupplement	
InCJKStrokes	
InCJKSymbolsandPunctuation	

InCJKUnifiedIdeographs	en:List of CJK Unified Ideographs
InCJKUnifiedIdeographsExtensionA	
InCombiningDiacriticalMarks	
InCombiningDiacriticalMarksforSymbols	
InCombiningDiacriticalMarksSupplement	
InCombiningHalfMarks	
InCombiningMarksforSymbols	
InControlPictures	
InCoptic	
InCurrencySymbols	
InCyrillic	Alfabeto cirílico
InCyrillicExtendedA	
InCyrillicExtendedB	
InCyrillicSupplement	en:List of Unicode characters#Cyrillic Supplement
InDevanagari	
InDingbats	
InEnclosedAlphanumerics	
InEnclosedCJKLettersandMonths	
InEthiopic	
InEthiopicExtended	
InEthiopicSupplement	
InGeneralPunctuation	
InGeometricShapes	
InGeorgian	

InGeorgianSupplement	
InGlagolitic	
InGreek	
InGreekExtended	
InGreekandCoptic	en:List of Unicode characters#Greek and Coptic
InGujarati	
InGurmukhi	
InHalfwidthandFullwidthForms	
InHangulCompatibilityJamo	
InHangulJamo	
InHangulSyllables	
InHanunoo	
InHebrew	en:List of Unicode characters#Hebrew
InHighPrivateUseSurrogates	
InHighSurrogates	
InHiragana	
InIdeographicDescriptionCharacters	
InIPAExtensions	en:IPA Extensions unicode block
InIdeographicDescriptionCharacters	
InKanbun	
InKangxiRadicals	
InKannada	
InKatakana	
InKatakanaPhoneticExtensions	

InKayahLi	
InKhmer	
InKhmerSymbols	
InLao	
InLatin-1 Supplement	en:Latin-1 Supplement unicode block
InLatinExtendedA	en:Latin Extended-A unicode block
InLatinExtendedB	en:Latin Extended-B unicode block
InLatinExtendedC	
InLatinExtendedD	
InLatinExtendedAdditional	
InLepcha	
InLetterlikeSymbols	
InLimbu	
InLowSurrogates	
InMalayalam	
InMathematicalOperators	
InMiscellaneousMathematicalSymbolsA	
InMiscellaneousMathematicalSymbolsB	
InMiscellaneousSymbols	en:Miscellaneous Symbols
InMiscellaneousSymbolsandArrows	
InMiscellaneousTechnical	
InModifierToneLetters	
InMongolian	
InMyanmar	

InNewTaiLue	
InNko	
InNumberForms	
InOgham	
InOlChiki	
InOpticalCharacterRecognition	
InOriya	
InPhagspa	
InPhoneticExtensions	
InPhoneticExtensionsSupplement	
InPrivateUse	
InPrivateUseArea	
InRejang	
InRunic	
InSaurashtra	
InSinhala	
InSmallFormVariants	
InSpacingModifierLetters	en:List of Unicode characters#Spacing Modifier Letters
InSpecials	
InSundanese	
InSuperscriptsandSubscripts	
InSupplementalArrowsA	
InSupplementalArrowsB	
InSupplementalMathematicalOperators	

InSupplementalPunctuation	
InSylotiNagri	
InSyriac	
InTagalog	
InTagbanwa	
InTaiLe	
InTamil	
InTelugu	
InThaana	
InThai	
InTibetan	
InTifinagh	
InUnifiedCanadianAboriginalSyllabics	
InVai	
InVariationSelectors	
InVerticalForms	
InYijingHexagramSymbols	
InYiRadicals	
InYiSyllables	en:List of Unicode characters in the Yi syllables block

Testar expressões regulares

Se encontrarem sites ou programas úteis coloquem aqui. De preferência não comerciais, sob uma licença livre e em português.

Sítios (on-line)

Site	Língua	Licença	Requisitos	Funcionalidades	Tem "ajudas"?	ExpReg compatíveis? ³
http://www.gskinner.com/RegExr/	Inglês		Java, Flash	☆☆☆	☆☆☆	✓
http://www.regex.com.br	Português e Inglês	BY-NC-SA 3.0	Java	☆☆☆	☆☆☆	✗
http://regexpal.com	Inglês	LGPL	Java	☆☆☆	☆☆☆	✗
http://regex.larsolavtorvik.com/#pcre	Inglês		Java	☆☆☆	☆☆☆	✓
http://regex.powertoy.org	Inglês	GPL	Java	☆☆☆	☆☆☆	
http://cuneytyilmaz.com/prog/jrx/	Inglês	GPL	Java	☆☆☆	☆☆☆	✗
http://www.nregex.com/nregex/	Inglês		Java	☆☆☆	☆☆☆	✗
http://www.myregextester.com	Inglês		Java	☆☆☆	☆☆☆	✓
http://tools.lymas.com.br/regexp_br.php	Português		Java	☆☆☆	☆☆☆	✗
http://www.idsklijnsma.nl/regexps/	Inglês		Java	☆☆☆	☆☆☆	✓
http://www.regextester.com	Inglês	CC 2.5	Java	☆☆☆	☆☆☆	

Programas

Nome	Plataforma	Língua	Licença	Funcionalidades	Tem "ajudas"? ²	ExpReg compatíveis? ³	Aceita texto de entrada longo? ⁴	Principal vantagem
https://addons.mozilla.org/en-US/firefox/addon/2077	n/a (<i>Firefox</i>)	Português (parcialmente)	GPL v2	★ ★ ★	★ ★ ★	✗		Extra (<i>addon</i>) para o Firefox
Regulator (http://sourceforge.net/projects/regulator/)	Windows	Inglês	LGPL	★ ★ ★	✓	✗	✓ ⁵	
regexbuilder (http://sourceforge.net/projects/regexbuilder/)	Windows	Inglês	GPL	★ ★ ★	✗	✗	✗	Bastante simples
reggy (http://code.google.com/p/reggy/)	Mac OSX	Inglês	GPL v2					
jregexptester (http://sourceforge.net/projects/jregexptester/)			LGPL v2.1					

Notas

- ¹ ↑ ***^a ^b*** Os metacaracteres ^ e \$ são adicionados automaticamente em MediaWiki:Titleblacklist - ver mw:Extension:SpamBlacklist#Syntax
- ² ↑ Não é necessário utilizar (*?i: expressão*) em MediaWiki:Titleblacklist - ver mw:Extension:Title Blacklist
- ³ ↑ ***^a ^b*** "ExpReg compatíveis?" - testado com a expressão regular \x{25} **Nota: Apesar da maioria dos programas e sítios não terem expressões regulares compatíveis com as utilizadas no MediaWiki, estes podem ser utilizados com as expressões regulares mais comuns.**
- ⁴ ↑ "Aceita texto de entrada longo" - programas testados com todos os títulos de páginas do domínio principal da Wikipédia lusófona (15.000.000 caracteres; 885.000 linhas)
- ⁵ ↑ O programa "Regulator" aceita um texto de entrada bastante longo mas a importação do texto é demorada (~15min). Após importar o texto, o programa fica um pouco lento apesar de ser razoavelmente rápido a detetar as correspondências das expressões regulares no texto.

Ver também

- Ajuda:Caracteres de escape (em HTML)
- Caractere de escape

- Sequência de escape

Ligações externas

- Tabelas Unicode no site oficial do Unicode Consortium (<http://www.unicode.org/charts/>) (em inglês) - disponíveis apenas no formato PDF
- Expressões regulares e o uso de Unicode (<http://unicode.org/unicode/reports/tr18/>) (em inglês)

Obtida de "http://pt.wikipedia.org/w/index.php?title=Ajuda:Expressões_regulares&oldid=36839576"

Categorias: !Guia de edição | !Ajuda

-
- Esta página foi modificada pela última vez à(s) 01h04min de 6 de setembro de 2013.
 - Este texto é disponibilizado nos termos da licença Atribuição-Partilha nos Mesmos Termos 3.0 não Adaptada (CC BY-SA 3.0); pode estar sujeito a condições adicionais. Consulte as condições de uso para mais detalhes.